Online Streaming

Adam Lyon (Fermilab)

October 2002 Collaboration Meeting

Outline:

- What is streaming, why is it important, why we need to do it?
- The online streaming plan; why do it this way?
- What does streaming mean for my analysis?
- Where is streaming now?

People involved & History

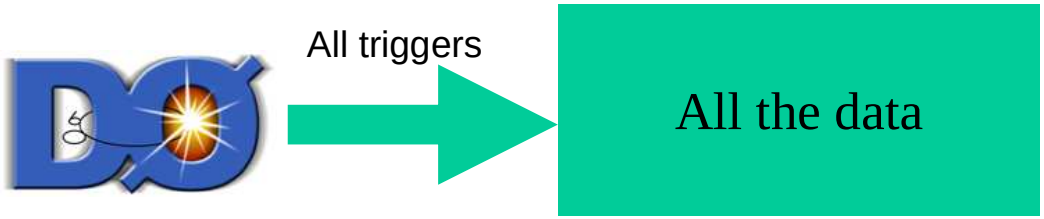
Michael Begel, Amber Boehnlein, Elizabeth Gallas, Stu Fuess, Jerry Guglielmo, Jonathon Hays, Greg Landsberg, Lee Lueking, Adam Lyon, Wyatt Merritt, Jianming Qian, Michelle Reschke, Heidi Schellman, Jeremy Simmons, (and the Analysis Tools Group), …

- Streaming from the Physics Groups:
  - Breese Quinn (Top), Rienhard Schwienhorst (Top), Su Yong Choi (Higgs), Peter Tamburello (Higgs), Marco Verzocchi (W/Z), Brad Abbott (B), Heriberto Castilla (B), Andrew Brandt (QCD), …

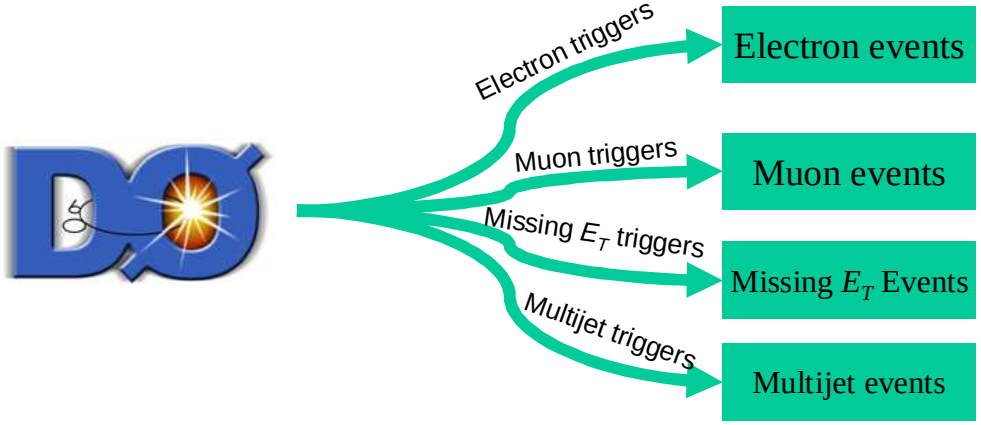
- Streaming has been discussed (lots) before:
  - D0Notes 3313 (1997), 3326 (1997), 3662 (1999)
  - +lots of documents on the web

What is streaming?

- *Splitting up raw data* ***online (at L3)*** *to different files*



- Often confused with **offline** production of derived samples
  - "stripping", "skimming", or "tapping out"



Nothing here prevents any offline stripping

Streaming

- Improves efficiency of data retrieval:
  - Events of interest are grouped in same files on same sets of tapes. Other events are sent elsewhere.
  - Minimize tape mounts for large samples *(only look at the tapes you care about)*
  - Minimize file access for large disk-resident samples *(only access the files you care about)*
  - Minimize I/O for running over large samples *(don't process events you don't care about)*

Streaming leads to more efficient use of resources

Why Stream Online?

- Improves efficiency online:
  - Better use of online disk (load balancing)
- Online Streaming most helpful for data on tape
- Reprocessing

  An ALL stream will be too large to re-reco. Streaming RAW files makes it easier to reprocess a subset and prioritize
  - New version of RECO
  - Improved calibration/alignment
  - "Expensive Options"
    - Different tracking program, lower thresholds
- Prioritized and Offsite Reco

  Different RAW streams will make reco'ing a subset of the data easier

Why Stream Online?

- Picking many events more efficient
  - If your desired events are evenly distributed, then picking from one stream means fewer tapes to mount

Why Stream Online?

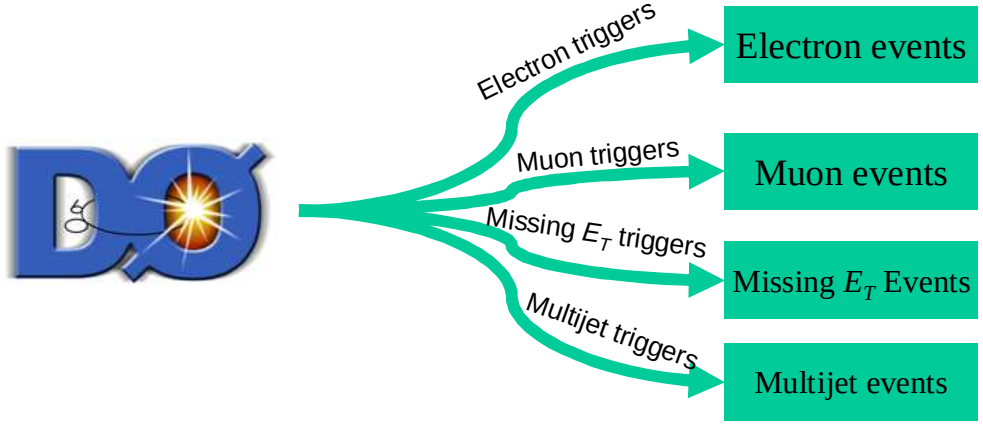
- Picking many events more efficient
  - If your desired events are evenly distributed, then picking from one stream means fewer tapes to mount

Why Stream Online?

- Streaming RAW propagates streams to all data-tiers for free
  - Though we imagine virtual-streaming for thumbnails on disk
- Online streaming immune to problems in RECO
  - Still sensitive to L3 trigger problems
  - Everyone is sensitive to L3 problems!
- Offline streaming in Run 1 was often far behind the first processing

Exclusive Streaming

- We can't afford the ideal picture of streaming:



- If an event satisfies two triggers (e.g. electron and jets)
  - In Run 1: event copied to all applicable streams (Inclusive streaming – 50% duplication rate in Run 1)
  - Can't do that in Run 2!
    - Tape costs are too high to have multiple copies of events
    - Multiple event copies eat online bandwidth
    - **Inclusive streaming overtaxes our limited resources!**

Exclusive Streaming

- Exclusive streaming for Run 2:
  - An event can to go one and only one stream
    - Except Mark and Pass events will go to a special inclusive `Monitor` stream (<5%; for trigger tests only, not used for physics)
  - Streaming performed by Level 3
  - Base streaming decision on **L3 triggered physics objects**

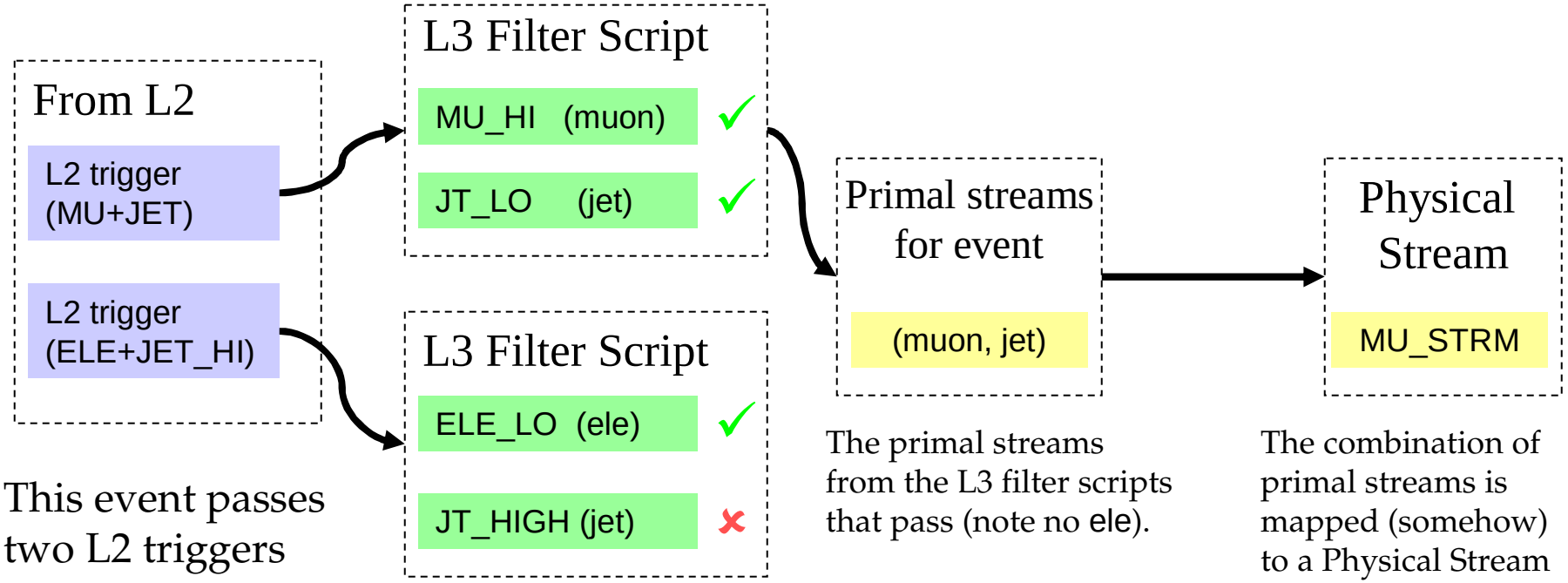
Streaming with L3 triggered physics objects

- Grouping like triggers:
  - L3 provides a natural way to do this
    - Keep track of the type of filters run for triggers that pass
    - There are filters that deal with electron ID, muon ID, jet ID, etc. -- Group together triggers that run these filters
    - Triggers that do not fall nicely into a group can be "overridden" to be in a certain group
  - Streaming decision by trigger group
    - A few groups (<10) is **much** more manageable than ~80 triggers
    - We call these groups **primal streams** (yes, a poor name)

Primal Streams

- An event may have more than one primal stream
- All combinations of n primal streams lead to 2^n-1 physical streams (actual files on the online disk)
  - If n is reasonable (*eg.* 7) that's a lot of streams (127)!
- Use a **Tree Algorithm** to reduce 2^n-1 to a much smaller number of **physical streams**
- Streaming with L3 triggered physics objects is ***equivalent*** to streaming by trigger bits
  - For any stream scheme based on primal streams and a tree, one can work backwards and make the equivalent trigger bit pattern → physical stream table **(no ambiguities)**

The Streaming Process: An example



This event passes two L2 triggers

Each L3 filter is associated with a "primal stream", *i.e.* the type of physics object it filters on [note two jet primals].
Pretend the event fails the lower script.

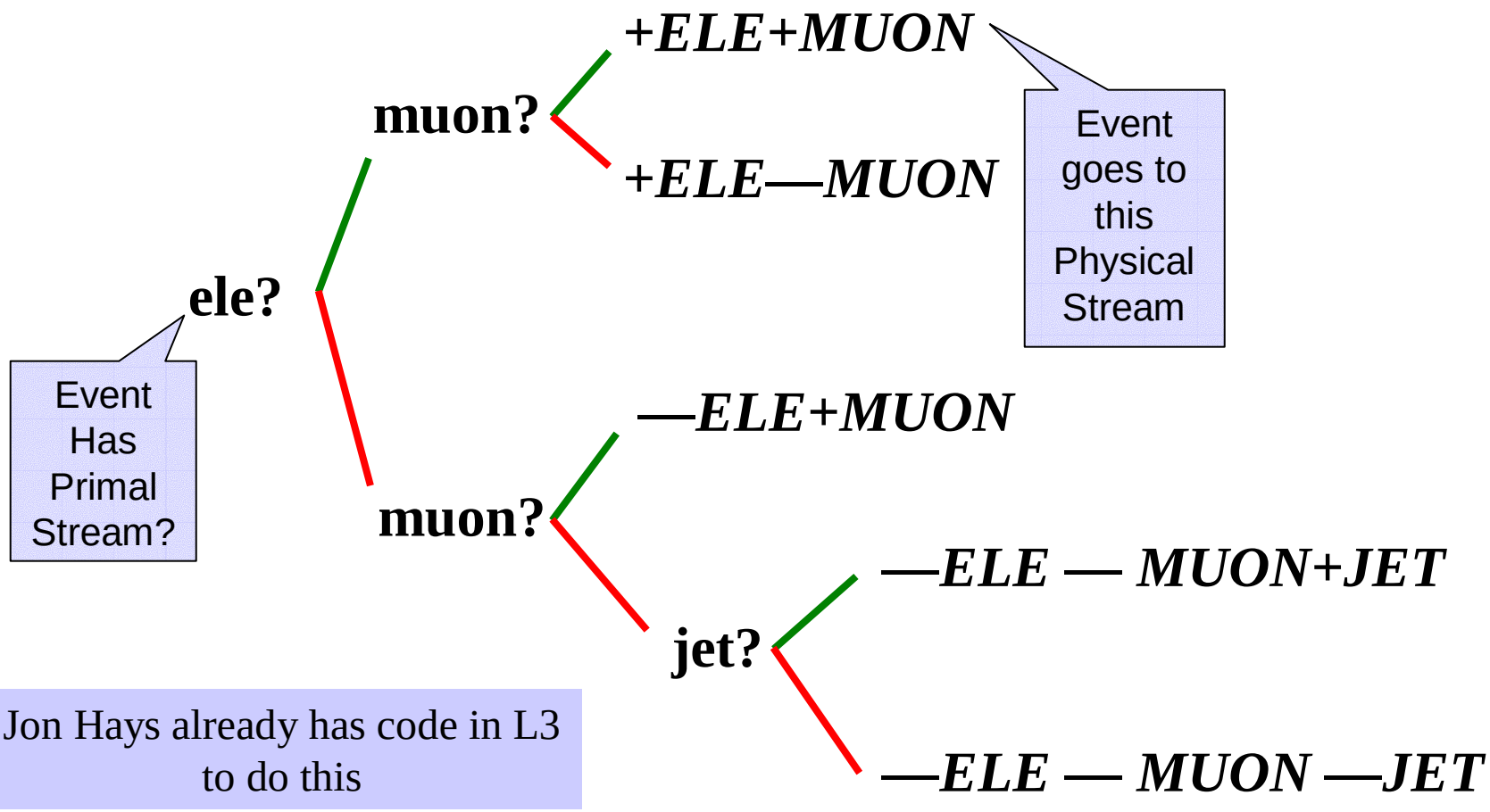
The primal streams from the L3 filter scripts that pass (note no ele).

If there are n primal streams, there are 2^n-1 possible combinations. This is probably too many.

The combination of primal streams is mapped (somehow) to a Physical Stream (the set of files on tape/disk).

We want m physical streams where $m << 2^n$-1

Tree Algorithm (an example)

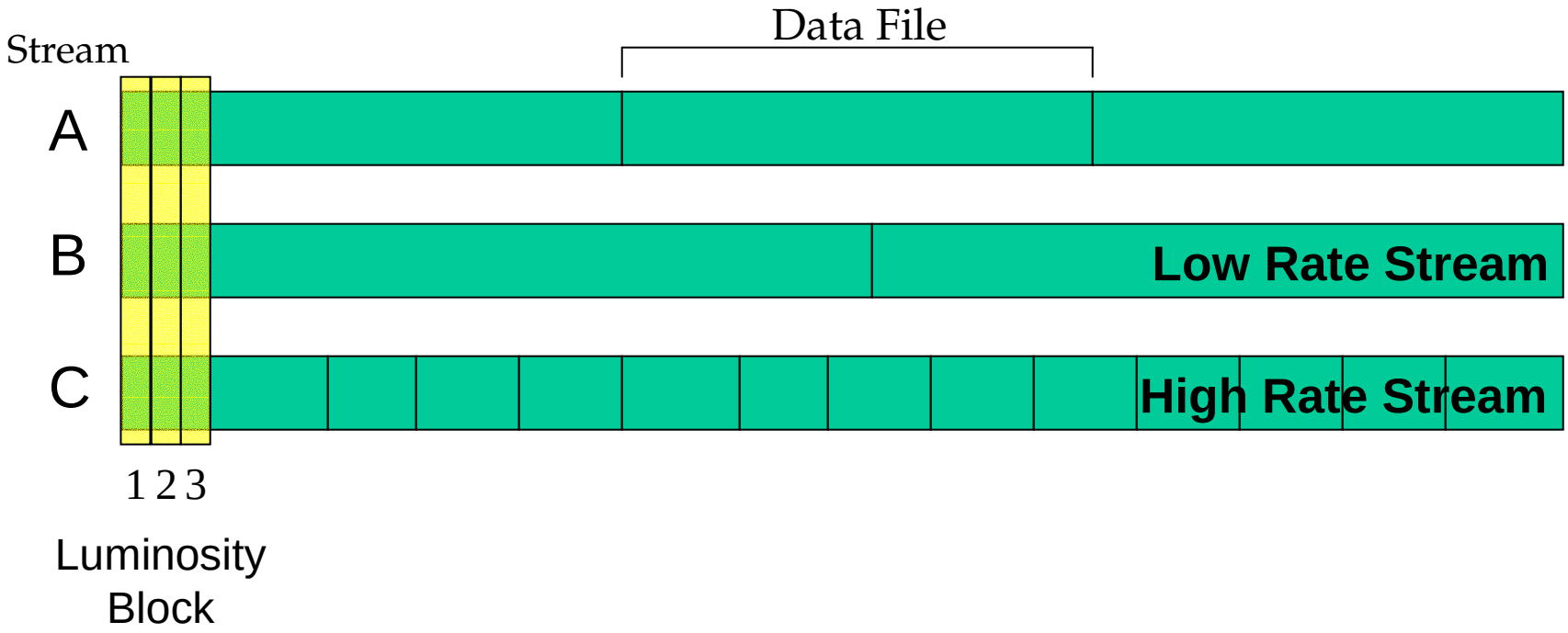


Jon Hays already has code in L3 to do this

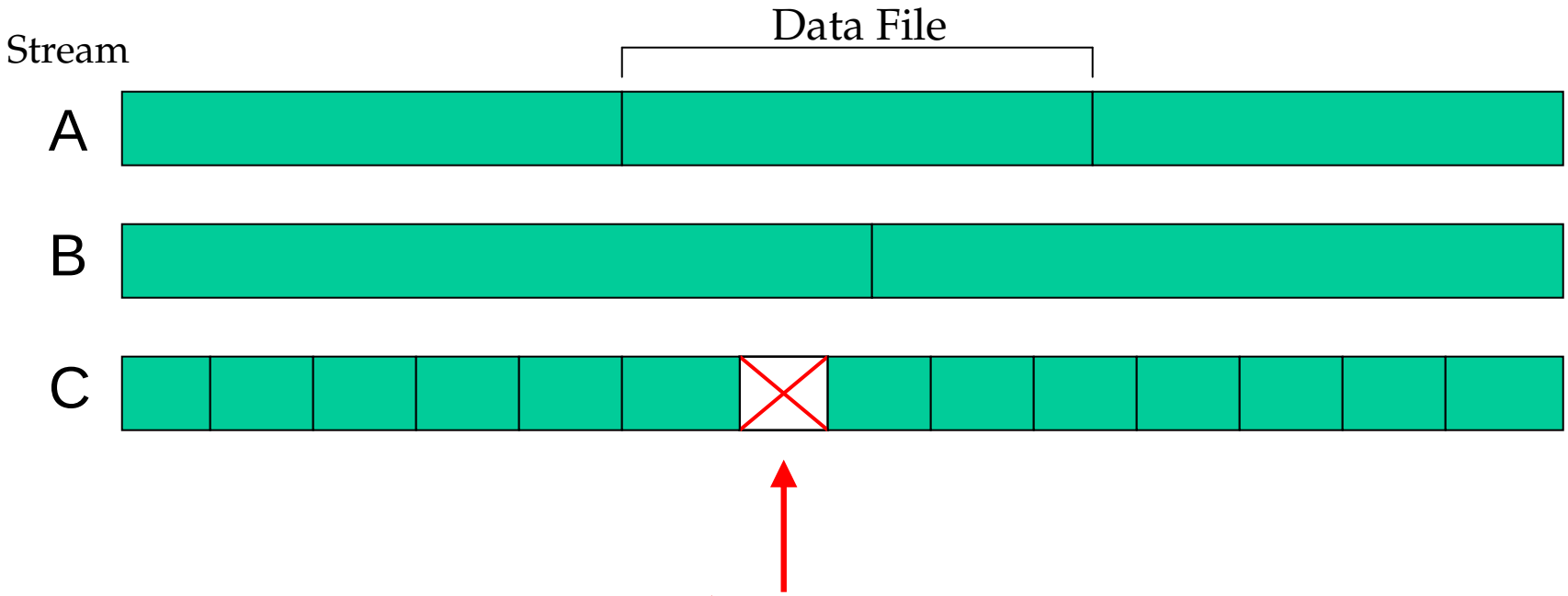
Streaming your analysis

- Your data may be in more than one stream
  - If an event with your trigger comes in with another trigger, the event may be "diverted" to a higher priority stream
  - If this is a rare occurrence or you veto, perhaps you don't care
  - Streams are topologically biased (of course)
- Most analyses will benefit -- process less data
  - And will have access to reprocessed data
  - But a few may have to process all the data
- SAM will make data access easy
  - Request your data by asking for trigger(s)
- Bookkeeping is more complicated
  - May be sensitive to missing data in other streams
  - Updating Automated bookkeeping system to deal with this…

Consequences for Bookkeeping



Consequences for Bookkeeping

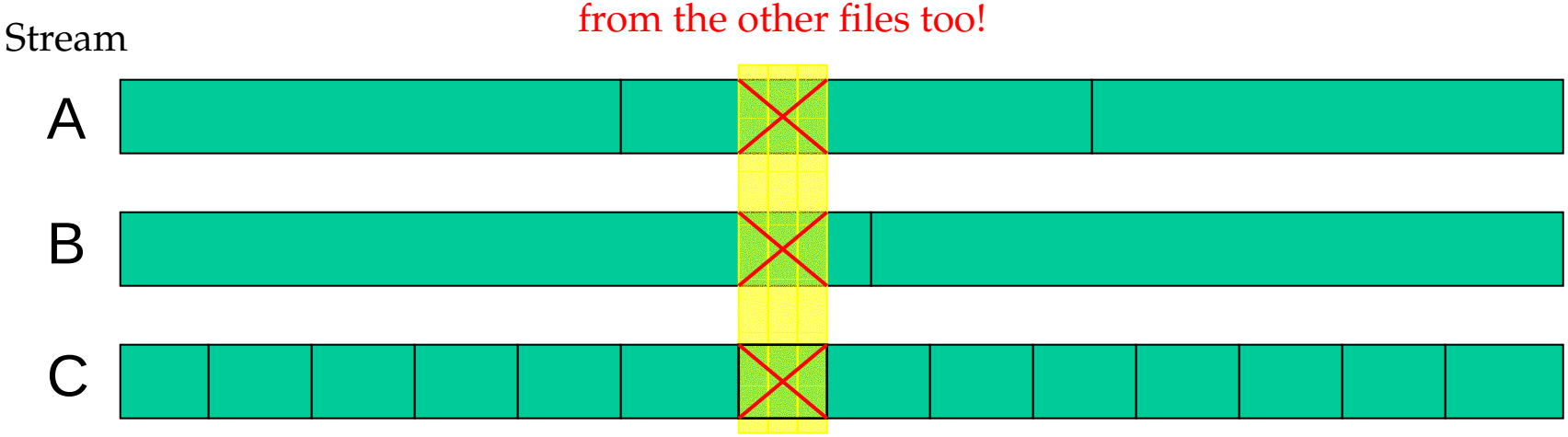


Let's say we can't use this file

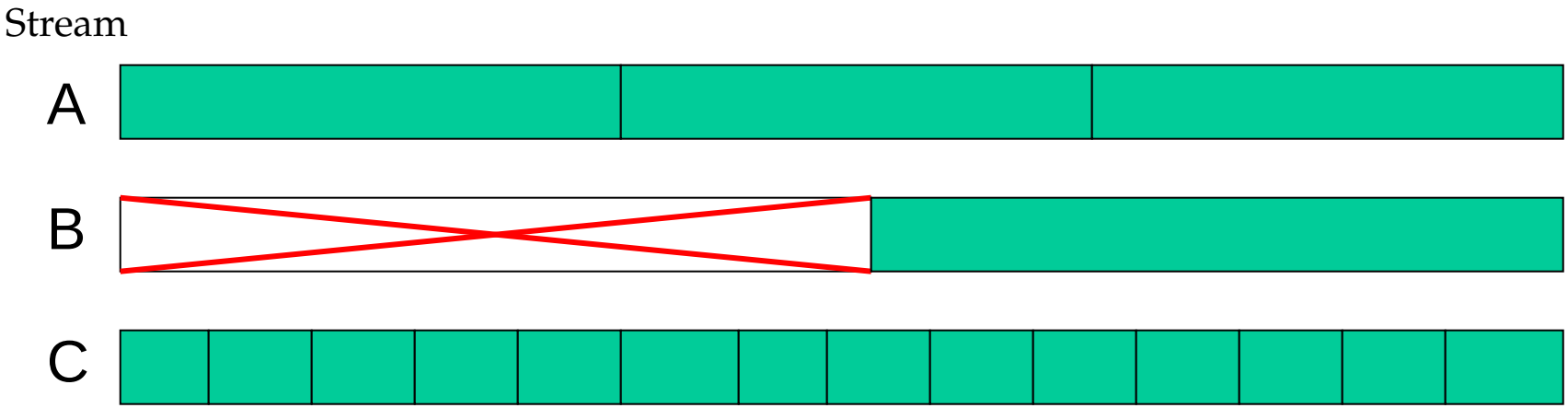
- File never made it to Reco
- Reco crashed on this file
- Sam couldn't deliver this file
- …

Consequences for Bookkeeping

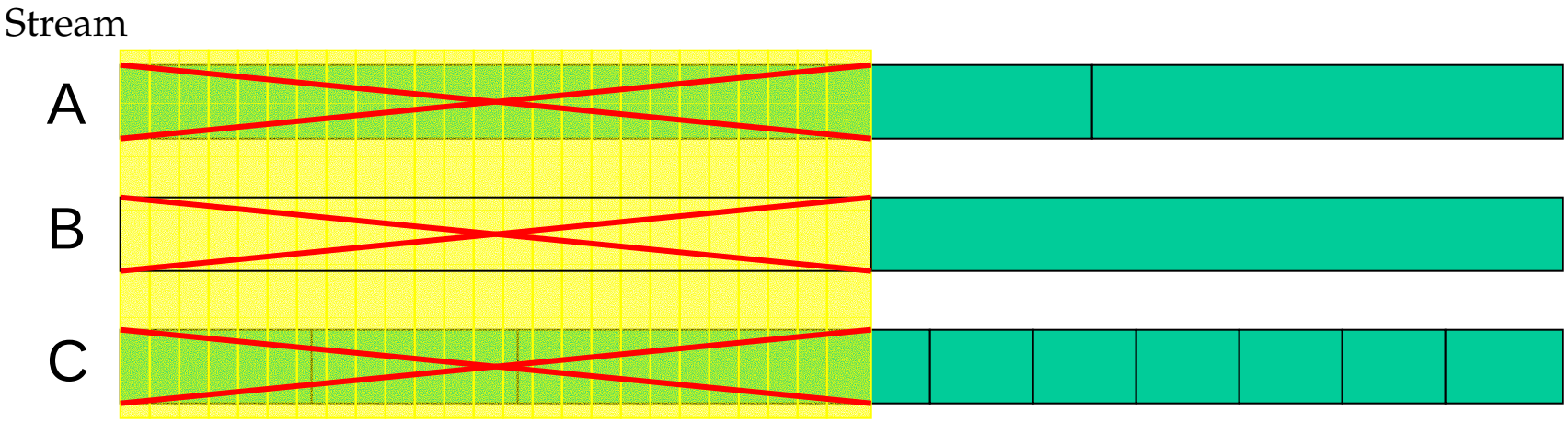
You may have to exclude
events and luminosity
from the other files too!



Consequences for Bookkeeping



Consequences for Bookkeeping



- Bad luminosity blocks in "rare" streams could lead to big subtractions
- Solution: Try not to have rare streams

How do we choose the Stream Scheme?

- This is the hard part
  - Need input from physics groups
  - The **A**nalysis **T**ools **G**roup has been driving this effort, until…

- The Plan
  - Two-stream streaming in a few days
  - Complex streaming in January

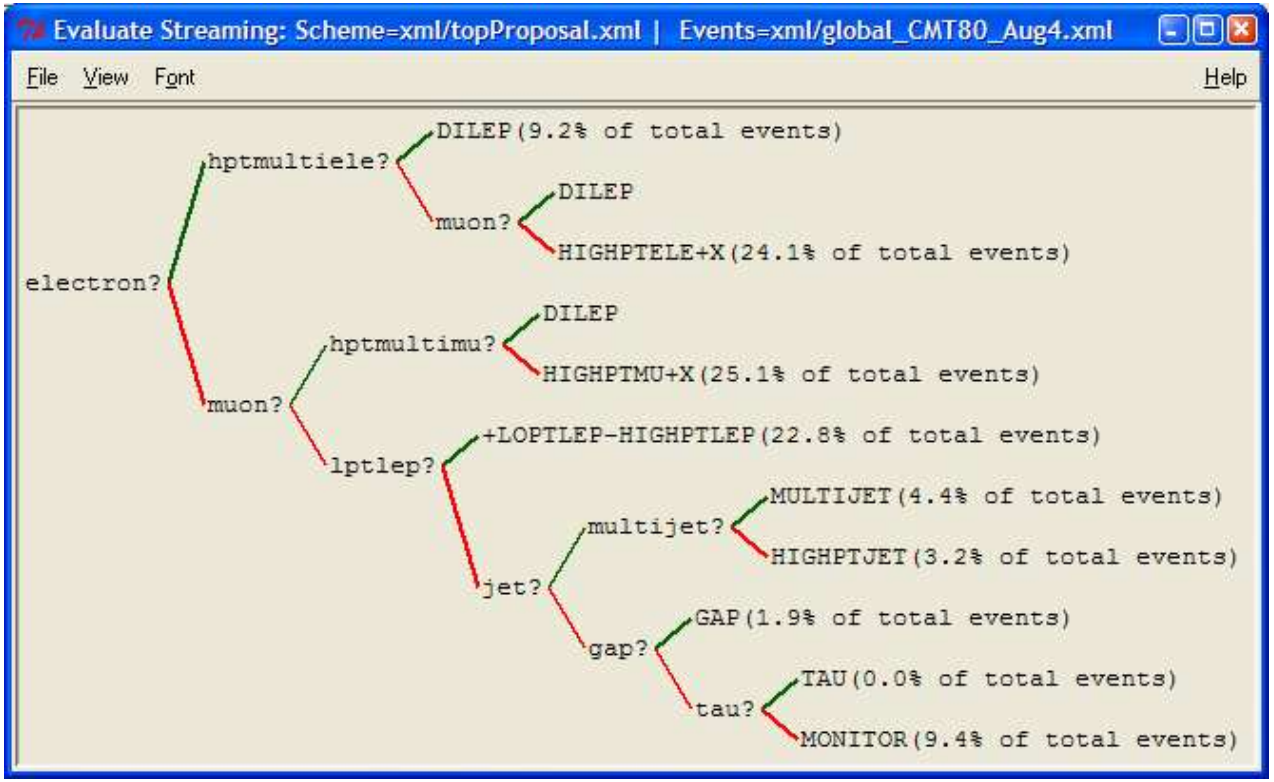
Streaming Plan

- Two-stream streaming
  - Proposed by Boaz (see his talks)
  - Shake down the streaming system
  - Provide ability to reprocess some data with p13 for Moriond
- Complex streaming
  - Several streams (~5)
  - Streams separated by physics (e.g. high p_T leptons, low p_T leptons, jets, separate e and μ, dilepton stream, …)

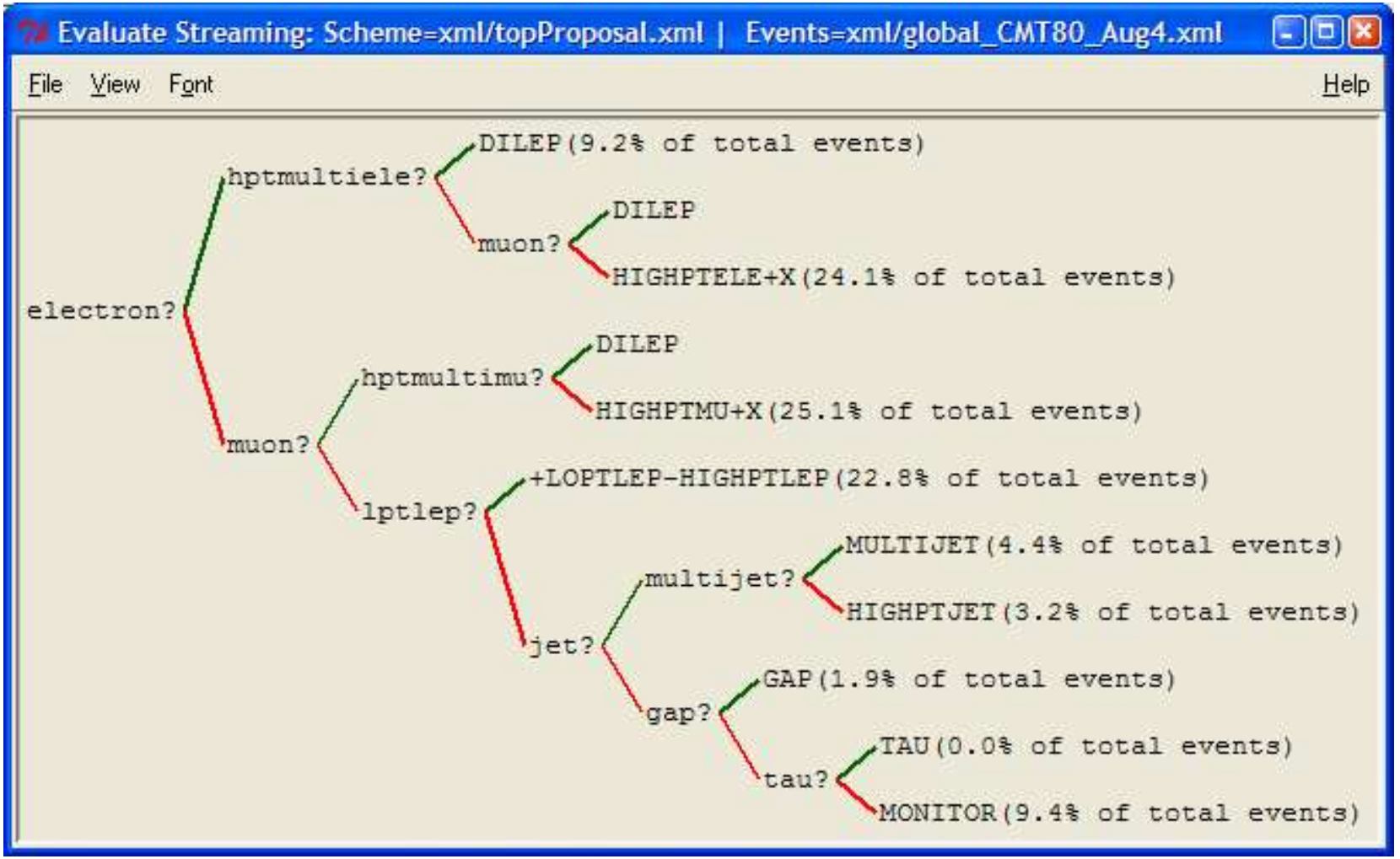
Tools to investigate streaming

evaluateStreaming:

- A streaming simulation tool
- Runs off of real data (SAM catalogs)
- Gives reports for sets of triggers:
  - Fraction of all data having your triggers
  - Fraction of data in needed streams
  - Fraction of data that's interesting in streams



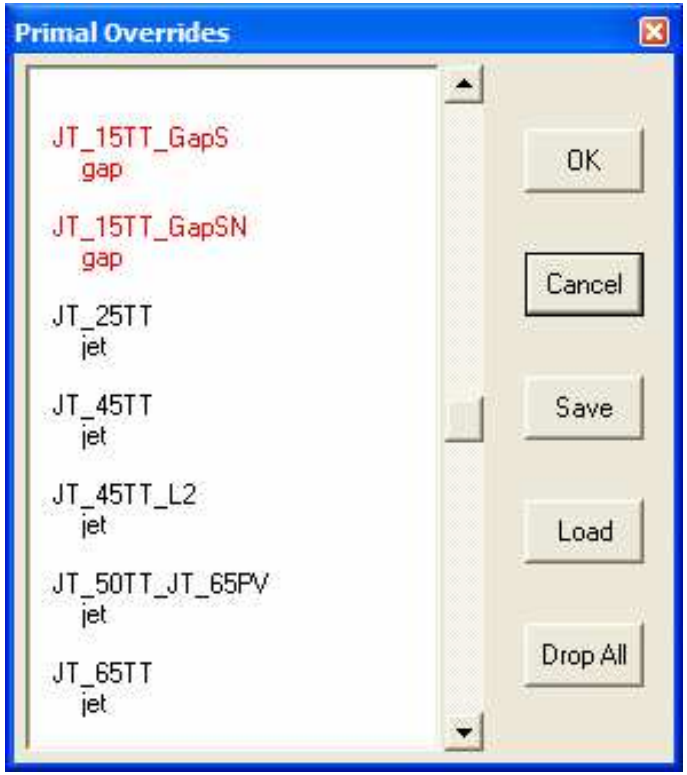
evaluateStreaming



Examples

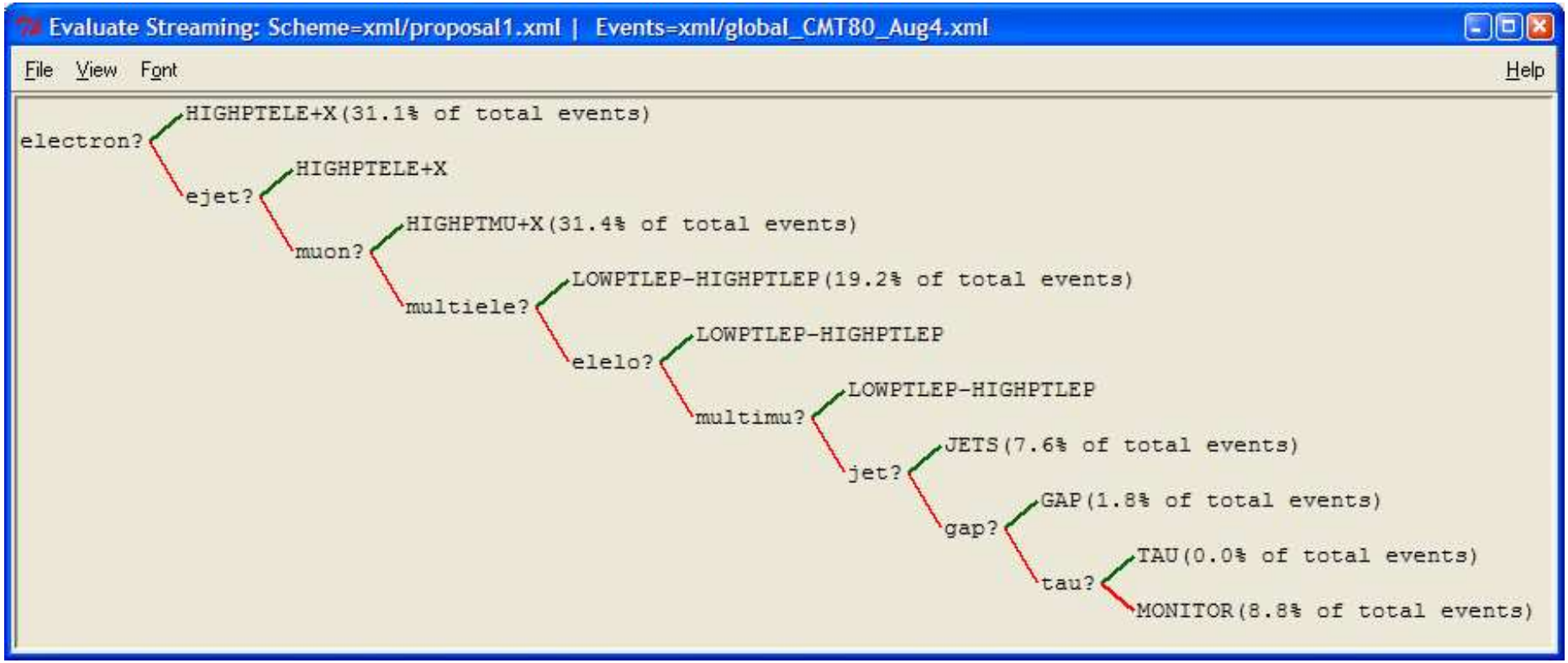
Example Trigger Report

```
Physical streams for trigger set: [EM_HI]
+-- Triggers represent 17045 events
    (2.1% of all 798520 events)
+-- 2 physical streams need to be processed:
    265539 events (33.3% of all events)
+-- 17045 events (6.4%) in these streams are
    interesting
+-- Breakdown of Physical Streams:
    +-- For Phys Stream HIGHPTELE+X: 14064 events
        of 192142 (7.3%) are interesting
    +-- For Phys Stream DILEP: 2981 events of
        73397 (4.1%) are interesting
```



Example Primal Stream Overrides

Another Proposal



Complex streaming - Current thinking

- High p_T leptons have the highest priority
  - W/Z analyses want as few streams as possible
  - Top/Higgs want similar streams as W/Z
- Debating finer grained streams
  - Some analyses will have fewer streams to look at
  - But perhaps more complicated bookkeeping, luminosity
- Realized that the choice of primal streams is key

Readiness

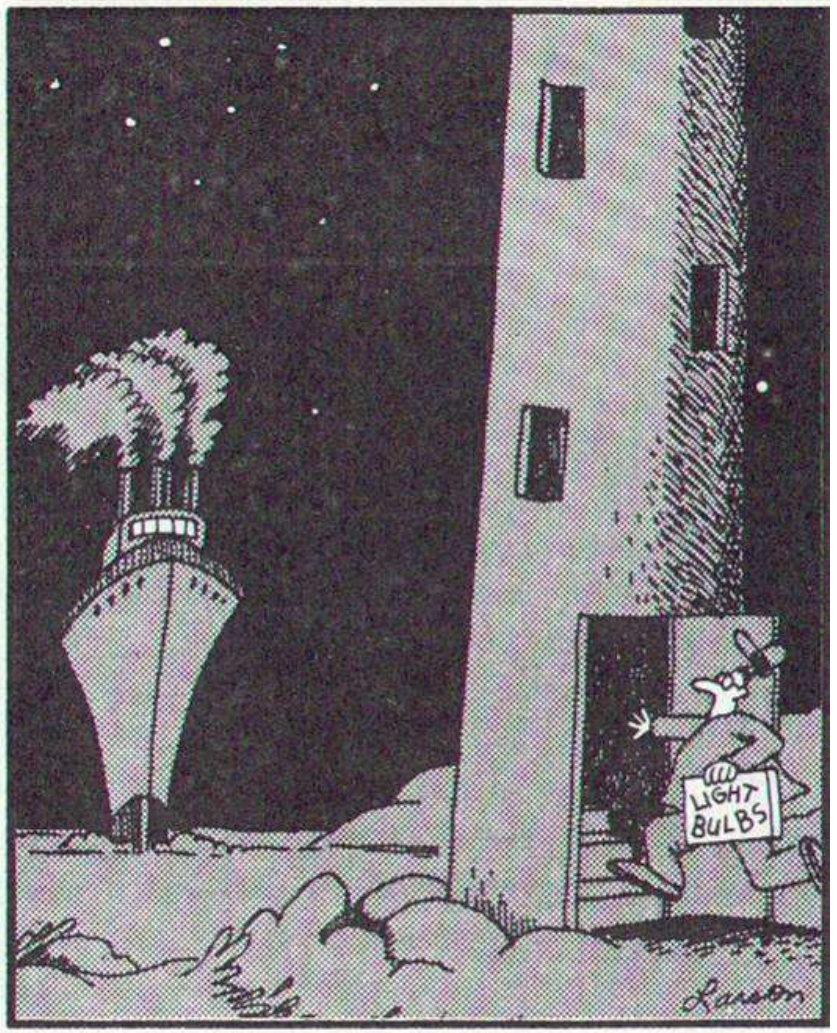
- Mechanically, streaming is *nearly* ready to go
  - p12 for L3 farm is "streaming-ready"
    - But still waiting for small fix to L3 Supervisor
    - p12 needs to be certified
  - Trigger database ready for overrides
    - Will be added by hand until web interface is ready
  - Stream table is in place
    - Stream scheme version tracked in RUNS database
  - Want to do an online test before implementation

Readiness

- Luminosity & Bookkeeping:
  - Online can handle two streams (won't declare every luminosity block bad)
  - Offline system needs to be updated (not just for streaming)
- Data Access:
  - Two streams are easy with SAM
  - Complex streaming will require some new tools
- Schedule:
  - Two-stream streaming in a week
  - Complex Streaming in January (start looking at stream schemes again end of this month)

Conclusions

- To use our resources efficiently, we must stream
- To use our online and processing farms efficiently, we must stream online
- To avoid overtaxing our resources, we must exclusively stream
- Implementation of streaming will be careful and staged. Discussions are ongoing…



What can you do?

- Look at information – lots on the web
  - Analysis Tools Web page (meeting minutes & presentations)
    http://www-d0.fnal.gov/computing/analysis_tools
  - Evaluate Streaming Tool Web page
    http://www-d0.fnal.gov/computing/analysis_tools/tools/evaluateStreaming
- Try some schemes:
  - *evaulateStreaming* is available on **d0mino** and **clued0**
  - Can load it on your laptop too
  - Lots of documentation
- Come to the ATG meetings

  Tuesdays 1pm in the Racetrack (WH7X) [Videoconferencing available]
  Sign up to d0-atg mailing list